Xiang, G., Ju, X. , and Holt, P. (2010) *Automatic 3D facial model and texture reconstruction from range scans.*Lecture Notes in Computer Science, 6169. pp. 260-269. ISSN 0302-9743

http://eprints.gla.ac.uk/62899/

Deposited on: 1st May 2012

Automatic 3D Facial Model and Texture Reconstruction From Range Scans

Guofu Xiang, Xiangyang Ju, and Patrik O'B. Holt

Cognitive Engineering Research Group, School of Computing,
The Robert Gordon University, Aberdeen, AB25 1HG, UK
{prs.xiang,x.ju,p.holt}@rgu.ac.uk

Abstract. This paper presents a fully automatic approach to fitting a generic facial model to detailed range scans of human faces to reconstruct 3D facial models and textures with no manual intervention (such as specifying landmarks). A Scaling Iterative Closest Points (SICP) algorithm is introduced to compute the optimal rigid registrations between the generic model and the range scans with different sizes. And then a new template-fitting method, formulated in an optmization framework of minimizing the physically based elastic energy derived from thin shells, faithfully reconstructs the surfaces and the textures from the range scans and yields dense point correspondences across the reconstructed facial models. Finally, we demonstrate a facial expression transfer method to clone facial expressions from the generic model onto the reconstructed facial models by using the deformation transfer technique.

Key words: surface reconstruction, texture reconstruction, range scans, scaling iterative closest points (SICP) algorithm, template fitting, expression transfer

1 Introduction

Modeling and animating realistic facial models is a substantial challenge in computer graphics, especially for facial expressions, because we are so familiar with human faces and very sensitive to "unnatural" subtle changes in faces. Such a challenge has drawn intensive academic and industrial interest in this area [8, 14]. However, creating a convincing synthetic character requires a tremendous amount of artistry and manual work. There is a clear need for more automatic techniques to reduce the painstaking work of artists and to make reuse of existing data.

One avenue for creating realistic facial models is 3D scanning technology. However, starting from a range scan, substantial effort is needed to process the noisy and incomplete surface into a model suitable for analysis and animation. Template-fitting methods are widely used for this purpose to fill holes, reduce the noise level, and capture characteristic features of range scans [1, 21, 22]. In addition, dense point correspondences, which are fundamental requirements in many applications such as morphing and shape analysis, can be also established across various models. Generally, template-fitting methods require users to provide a small set of manually specified landmarks to initially align or warp a template with targets [1, 21, 22]. The process of positioning landmarks seems to be tedious and error-prone.

Besides Modeling facial expressions directly from range scans of human faces, it would be better to reuse existing facial expressions to generate new ones on desired targets instead of creating them from scratch, which is the idea of *expression cloning* [12]. One key problem for expression cloning is to build good dense correspondences between models.

In this paper, we present a fully *automatic* approach to reconstructing 3D facial models and textures from range scans without requiring manual intervention. This paper makes several specific technical contributions. First, we introduce a *Scaling Iterative Closest Points* (SICP) algorithm to compute the optimal rigid registrations between a generic template facial model and range scans with different sizes. Second, we propose a unified optimization framework to reconstruct facial surfaces and textures from the range scans. We also present a method to automatically generate new facial expressions on the reconstructed facial models from expressions on the generic model by using the deformation transfer technique.

In the following section, we review some topics related to our work. In Section 3, we present the details of SICP to rigidly register a template facial model to range scans with different scales and show an optimization framework to reconstruct facial models and textures from range scans. Results and conclusions are presented in Sections 5 and 6, respectively.

2 Related Work

Modeling and synthesizing faces is an active research field in computer graphics and computer vision. Here we review three topics most related to our current work: ICP-based registration, template fitting, and expression transfer. Other related work is discussed throughout the paper, as appropriate.

ICP-based Registration Since the first paper of ICP [2], ICP has been widely used for geometric alignment of 3D models and many variants of ICP have been proposed [16]. Generally, the original ICP can only deal with models with the same scale. To account for the scale problem, Du *et al.* proposed an extension of the ICP algorithm, called the Iterative Closest Points with Bounded Scale (ICPBS) algorithm, which integrated a scale parameter with boundaries into the original one [6], but it's unclear how to determine the upper and lower boundaries of scales that contain the optimal scale.

Template Fitting Due to its great challenge in many research fields, numerous research efforts are devoted to establishing correspondences between different meshes [9]. The template-fitting method [1, 17] deforms a template to a target object to minimize the combining errors of smoothness and fitness between them. Recently, template fitting has become particular popular due to its simplicity and robustness to noisy range data [11, 21]. Our reconstruction method shares the similar idea, but it is derived from physically based elastic deformations of thin shells by variational methods [4].

Expression Transfer Noh and Neumann first proposed the concept of *expression cloning* that facial expressions of one 3D facial model were copied onto other facial models [12]. The dense point correspondences were established by volume morphing with Radial Basis Functions (RBFs) through dozens of initial corresponding points. Sumner *et. al.* [5, 17] generalized the idea to transfer arbitrary nonlinear deformation exhibited by a source triangle mesh onto different target triangle meshes. To build triangle correspondences, they manually specified a small set of initial corresponding feature points and then fitted the source meshes to the target using the template-fitting method. Vlasic *et al.* proposed a method, which used multilinear models for mapping video-recorded performance of one individual to facial animations of another [20]. An example-based approach [15] proposed by Pyun *at al.* clones facial expressions of a source model to a target model while reflecting the characteristic feature of the target model.

3 Automatic Facial Model and Texture Reconstruction

In this paper, we assumed that the range scans to reconstruct were upright front faces, in which some other unwanted parts (such as hair, neck, shoulder) might also present. Given such a range scan, our goal is to build a new facial model with texture to reflect the shape and texture of the range scan from a template facial model. The missing data in the facial region of the range scan should be filled and the noise level should be reduced as well.

Our reconstruction method consists of two steps: the first step is to compute the initial rigid registration between a template and a range scan; the second step is to iteratively deform the template model toward the range scan to capture the shape of the range scan and the texture is obtained in the same way.

We prefer triangle meshes for the representation of our models and range scans for efficiency and simplicity. Before elaborating our method, let us introduce some notations used in this paper. A triangle mesh $\mathcal{M}$ consists of a geometrical and a topological component, i.e., $\mathcal{M} = (\mathcal{P}, \mathcal{K})$, where the latter can be represented by a simplicial complex with a set of vertices $\mathcal{V} = \{v_i, 1 \leq i \leq |\mathcal{V}|\}$[1], edges $\mathcal{E} = \{e_i \in \mathcal{V} \times \mathcal{V}, 1 \leq i \leq |\mathcal{E}|\}$ and triangles $\mathcal{F} = \{f_i \in \mathcal{V} \times \mathcal{V} \times \mathcal{V}, 1 \leq i \leq |\mathcal{F}|\}$. The geometric embedding of a triangle mesh into $\mathbb{R}^3$ is specified by associating a 3D position $\mathbf{p}_i$ for each vertex $v_i \in \mathcal{V}$: $\mathcal{P} = \{\mathbf{p}_i := \mathbf{p}(v_i) \in \mathbb{R}^3, 1 \leq i \leq |\mathcal{V}|\}$.

3.1 SICP Registration

In order to reconstruct the surface of a range scan using a template, we need first roughly place the template close to the range scan. Traditionally, this is

[1] $|\cdot|$ denotes the number of elements in the set.

done by manually specifying a small set of landmarks [1, 17, 21, 22]. Our method deals with this problem with no manual intervention.

[?] Since the template facial model and the range scans of human faces have much similarity in shape, it is intuitive to use the ICP algorithm to compute the initial rigid registrations between them. However, there is a challenge dealing with the scale problem, because the size of the facial region in the range scans is not known a priori and the range scans may also include some unwanted parts (see Figure 4).

To deal with the scale problem, we employed an extension version of the ICP algorithm, called the Scaling Iterative Closest Points (SICP) algorithm [7], which integrates a scale parameter s to the original ICP equation and iteratively refines the scale from an estimated initial scale until convergence.

Given a template mesh $\mathcal{M}_{\text{template}}$ and a range scan mesh $\mathcal{M}_{\text{scan}}$, the goal of SICP is to find the transformation (scale s, rotation $\mathbf{R} \in \mathbb{R}^{3\times 3}$ and translation $\mathbf{t} \in \mathbb{R}^3$) so that the distance between the registered template mesh $\mathcal{M}'_{\text{template}}$ and $\mathcal{M}_{\text{scan}}$ is as close as possible. Obviously, we should avoid degenerate cases such as $s = 0$ by providing a good initial value for s.

As the original ICP algorithm, SICP is an iterative algorithm, which iteratively refines the registration based on previous registrations until it satisfies a certain termination condition. Let us denote a sequence of registrations by $\mathcal{T} = \{\mathbf{T}_k = (s_k, \mathbf{R}_k, \mathbf{t}_k), 0 \le k \le |\mathcal{T}|\}$. Then the registration process can be formulated mathematically as follows,

$$\mathcal{C}_{k+1} = \{\arg\min\nolimits_{c \in \mathcal{M}_{\text{scan}}} d(s_k \mathbf{R}_k \mathbf{p}_i + \mathbf{t}_k, \mathbf{c})\}\,, \tag{1}$$

$$(s_{k+1}, \mathbf{R}_{k+1}, \mathbf{t}_{k+1}) = \arg\min\nolimits_{s,\mathbf{R},\mathbf{t}} \sum_{i=1}^{|\mathcal{P}_{\text{template}}|} \|s\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{c}_i\|^2, \mathbf{c}_i \in \mathcal{C}_k\,, \tag{2}$$

where $\mathbf{p}_i \in \mathcal{M}_{\text{template}}$, $d(\cdot)$ is a distance function. Equation 1 is to find the corresponding closest points on $\mathcal{M}_{\text{scan}}$ for the points of $\mathcal{M}_{\text{template}}$ and Equation 2 is the absolute orientation problem [10].

As mentioned above, the initial registration state, $s_0, \mathbf{R}_0, \mathbf{t}_0$, is important for the local convergence of SICP. In our examples, we set the initial values as following,

$$s_0 = \frac{\sum_{i=0}^{N} |\mathbf{q}_i - \bar{\mathbf{q}}|/N}{\sum_{i=0}^{M} |\mathbf{p}_i - \bar{\mathbf{p}}|/M}, \quad \mathbf{R}_0 = \mathbf{I}, \quad \mathbf{t}_0 = \bar{\mathbf{q}} - s_0 \mathbf{R}_0 \bar{\mathbf{p}}\,, \tag{3}$$

where $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$ are the centroids of the template and the scan meshes, M and N the number of points of the two meshes, and $\mathbf{I}$ the 3×3 identity matrix. Although SICP has many degenerate cases and does not guarantee the global convergence, our tests show its capability to register the template to different range scans (see Figures 1 and 4).

3.2 Deformable Model

Due to the shape diversities between the template facial model and range scans, we need further deform the template after the initial rigid registration. There are two criteria that should be considered during the deformation process. One is the regularity that penalizes dramatic changes in mesh. Another criterion is the fitting error, which can be formulated as the total distance between corresponding points.

Since the template mesh is a two-manifold surface, the change of the surface can be measured by the change of the first and the second fundamental forms and therefore yields a measure of stretching and bending [18]. Given a two-manifold surface $\mathcal{S}$, after deformation, it becomes $\mathcal{S}'$, we can represent the deformed surface $\mathcal{S}'$ by $\mathbf{p}' = \mathbf{p} + \mathbf{d}$, where $\mathbf{p} \in \mathcal{S}$, $\mathbf{p}' \in \mathcal{S}'$, and $\mathbf{d}$ is the displacement. The minimization of the physically based elastic energies yields the so-called Euler-Lagrange partial differential equation (PDE) [4]:

$$-k_s \Delta \mathbf{d} + k_b \Delta^2 \mathbf{d} = 0\,, \tag{4}$$

where k_s and k_b are coefficients, Δ and Δ^2 represent the Laplacian and the bi-Laplacian operator, respectively. The Laplacian operator can be extended to triangle meshes to obtain the discrete form of the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ (refer to [4]). Thus, we can formulate our deformable model as follows,

$$\min_{\mathbf{d}_i} \sum_{i=1}^{M} \| - k_s \Delta_{\mathcal{M}_{\text{template}}} \mathbf{d}_i + k_b \Delta^2_{\mathcal{M}_{\text{template}}} \mathbf{d}_i \|^2 + k_c \sum_{i=1}^{M} w_i \|\mathbf{d}_i - (\mathbf{c}_i - \mathbf{p}_i)\|^2, \tag{5}$$

where $\mathbf{p}_i \in \mathcal{P}_{\text{template}}$, $\mathbf{c}_i \in \mathcal{M}_{\text{scan}}$ is the corresponding closest point of $\mathbf{p}_i$, $\mathbf{d}_i$ is the unknown displacement, and k_s, k_b, k_c represent the contribution of stretching, bending and fitting in the total energy, respectively. $w_i = 1$ if the corresponding closest point satisfies a certain compatible conditions, otherwise 0. We employed the similar compatible conditions as [17, 19] to reject pseudo point matching, such as, requiring the angle between two corresponding normals should be greater than 60 degrees, rejecting boundary vertices. The minimization problem can be reformulated as a sparse linear system in terms of least squares [4].

An annealing-like deformation scheme is employed in our experiments. At the initial stage, k_s and k_b are set to relatively large values compared to k_c (In our tests, k_s, k_b and k_c are initially set to 50, 20, 2, respectively). Because at the initial stage we cannot estimate good correspondences between the template and the range scan by the closest points due to the shape diversity and large values of k_s and k_b do not allow dramatic change of the mesh. Then we relax the stiffness of the template facial model by gradually decreasing the values of k_s and k_b toward 1.

3.3 Texture Reconstruction

Texture can improve the reality of facial models. Thus it is desirable to make the textures available for the reconstructed facial models. However, the original range scans usually have holes (missing data). We cannot find all the texture coordinates for the reconstructed facial models.

We solve the texture reconstruction problem in the similar way proposed in the previous section, but here we consider the texture coordinates $\mathbf{u}_i \in \mathbb{R}^2$ as the unknown variables and the equation becomes

$$\min_{\mathbf{u}_i} \sum_{i=1}^{M} \| - k_s \Delta_{\mathcal{M}_{\text{template}}} \mathbf{u}_i + k_u \Delta^2_{\mathcal{M}_{\text{template}}} \mathbf{u}_i \|^2 + k_c \sum_{i=1}^{M} w_i \|\mathbf{u}_i - \mathbf{u}'_i\|^2, \quad (6)$$

where $\mathbf{u}'_i$ is the texture coordinates of the corresponding closest point on the range scan for the point $\mathbf{p}_i$.

When reformulating Equations 5 and 6 in matrix form, we can see that the two equations have the same sparse matrix and only differ in the right hand side. Thus the texture reconstruction can be efficiently solved because the sparse matrix is only factorized once.

4 Facial Expression Transfer

After the facial model and texture reconstruction, all the reconstructed facial models have the same topology as the template one, i.e., the dense point correspondences are automatically established across models. These dense correspondences have numerous applications in many areas such as shape space analysis [1], linear facial model [3], morphing. In this paper, to demonstrate the reconstructed facial models, textures and the correspondences, we show the facial expression transfer from the generic facial model onto various reconstructed facial models by using the deformation transfer technique [17]. The results are shown in Figure 5.

5 Results

We reconstructed 3D facial models from six 3d range scans, which are from the Face Recognition Grand Challenge (FRGC ver2.0) data set [13]. The statistics for the results are shown in Table 1. All computations were performed on a 2.4 GHz Intel Core2 CPU machine with 3 GB RAM. Timings are measured

Table 1. Statistics for the results shown in Figure 4

ID	#Points	#Triangles	Registration Time	Reconstruction Time	Total Time
template	1880	3580	-	-	-
02463d550	104425	205176	38	51	89
04485d284	112154	221296	49	68	117
04202d438	60544	118766	28	48	76
04201d368	103061	202160	43	54	79
04213d280	120792	234534	34	53	87
04279d283	112497	219790	38	52	90

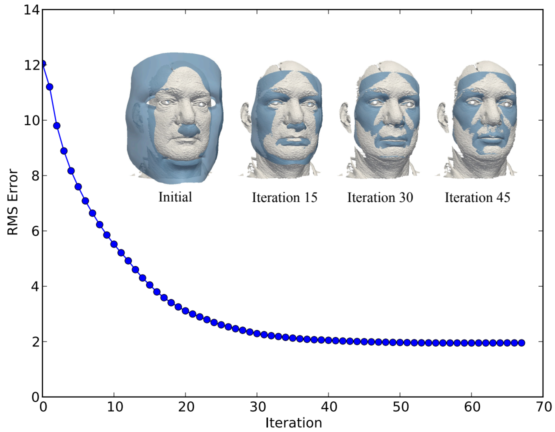

Fig. 1. The RMS error of SICP registration. The inset figures show the overlap between the template model and the range scan (02463d550) during the registration.

in seconds and exclude I/O operations. The order of the IDs of range scans in Table 1, which are the unique numbers in FRGC, is the same as that in Figure 4 (a) and (b).

Figure 1 shows the curve of the root-mean-squared (RMS) error during the SICP registration of the template to the range scan (02463d550). The curve definitely indicates the convergence of SICP, which is also shown by the inset figures.

Figure 2 shows the deformation process during reconstruction of the range scan (02463d550). The distances from the template to the range scan are encoded into colours. As we can show from the figure, the reconstruction error rapidly decreases across the face during the first several iterations.

To demonstrate the results of texture reconstruction, we rendered the range scans and reconstructed template facial model with a checkerboard texture and the original texture respectively as shown in Figure 3. We can see that the facial features are faithfully matched between the template and the range scan. The

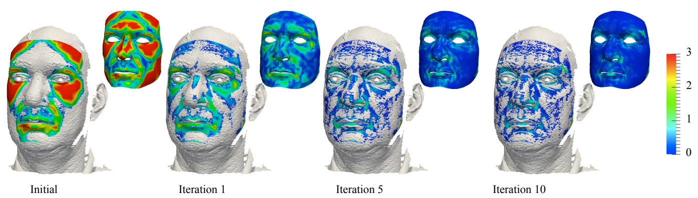

Fig. 2. Deformation process of the deformable model. The colour mapping shows the distances between the template and the range scan (02463d550).

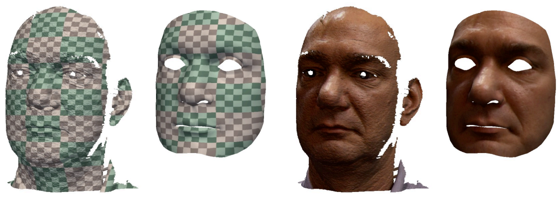

Fig. 3. The results of texture reconstruction.

reconstructed facial model along with the reconstructed texture (the rightmost in Figure 3) is more realistic than the original range scan as the holes are filled and the noise level is reduced.

We performed the facial expression transfer experiments of cloning five expressions from the template facial model onto three reconstructed facial models. The results are presented in Figure 5.

6 Conclusions and Future Work

We have presented a robust algorithm for 3D facial model and texture reconstruction from range scans of human faces. One of the main benefits of our method is fully automatic. Our method requires no manual intervention and we do not require a small set of corresponding feature landmarks. Our system demonstrates that high quality results can be obtained for a variety of range scans, with a realistic reconstruction of shape and texture. Key to the success of our algorithm is the robust rigid registration based on Scaling Iterative Closest Points (SICP) algorithm and the template fitting based on an elastic deformable model. As future work, we plan to extend our method to 4D range scans. We want to track a temporal sequence of range scans, faithfully reproduce the motion sequences in reconstructed facial models, and then transfer the motion sequences onto any other facial models.

Acknowledgments

This work is supported by the Robert Gordon University Research Development Initiative (RDI) PhD Studentship. The authors would like to thank the FRGC[13] organizers for providing data.

References

1. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. In: Proc. SIGGRAPH '03 Papers. pp. 587–594. ACM, San Diego, California, USA (2003)
2. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2), 239–256 (1992)

3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proc. SIGGRAPH '99. pp. 187–194. ACM, Los Angeles, California, USA (1999)
4. Botsch, M., Sorkine, O.: On linear variational surface deformation methods IEEE Trans. Vis. Comput. Graph. 14(1), 213–230 (2008)
5. Botsch, M., Sumner, R.W., Pauly, M., Gross, M.: Deformation transfer for detail-preserving surface editing. In: Proc. VMV '06. pp. 257–364. Aachen, Germany (2006)
6. Du, S., Zheng, N., Ying, S., Wei, J.: ICP with bounded scale for registration of m-D point sets. In: Proc. ICME '07. pp. 1291–1294. IEEE Computer Society, Beijing, China (2007)
7. Du, S., Zheng, N., Ying, S., You, Q., Wu, Y.: An extension of the ICP algorithm considering scale factor. In: Proc. ICIP '07. vol. 5, pp. 193–196. IEEE Computer Society, San Antonio, Texas, USA (2007)
8. Haber, J., Terzopoulos, D.: Facial modeling and animation. In: Proc. SIGGRAPH '04 Courses. p. 6. ACM, Los Angeles, California, USA (2004)
9. Hormann, K., Polthier, K., Sheffer, A.: Mesh parameterization: theory and practice. In: Proc. SIGGRAPH Asia '08 Courses. pp. 1–87. ACM, Singapore (2008)
10. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A 4(4), 629–642 (1987)
11. Li, H., Adams, B., Guibas, L.J., Pauly, M.: Robust single-view geometry and motion reconstruction. In: Proc. SIGGRAPH Asia '09 Papers. pp. 175:1–10. ACM, New York, NY, USA (2009)
12. Noh, J.y., Neumann, U.: Expression cloning. In: Proc. SIGGRAPH '01. pp. 277–288. ACM, Los Angeles, California, USA (2001)
13. Phillips, P.J., Flynn, P.J., Scruggs, W.T., Bowyer, K.W., Chang, J., Hoffman, K.J., Marques, J., Min, J., Worek, W.J.: Overview of the face recognition grand challenge. In: Proc. CVPR '05. pp. 947–954. IEEE Computer Society, San Diego, CA, USA (2005)
14. Pighin, F., Lewis, J.P.: Performance-driven facial animation. In: Proc. SIGGRAPH '06 Courses. ACM, San Diego, California, USA (2006)
15. Pyun, H., Kim, Y., Chae, W., Kang, H.W., Shin, S.Y.: An example-based approach for facial expression cloning. In: Proc. SCA '03. pp. 167–176. Eurographics Association, San Diego, California (2003)
16. Rusinkiewicz, S., Levoy, M.: Efficient variants of the icp algorithm. In: Proc. 3DIM '01. pp. 145–152. IEEE Computer Society, Quebec City, Canada (2001)
17. Sumner, R.W., Popović, J.: Deformation transfer for triangle meshes. In: Proc. SIGGRAPH '04 Papers. pp. 399–405. ACM, Los Angeles, California, USA (2004)
18. Terzopoulos, D., Platt, J., Barr, A., Fleischer, K.: Elastically deformable models. In: Proc. SIGGRAPH '87. pp. 205–214. ACM, Anaheim, California, USA (1987)
19. Turk, G., Levoy, M.: Zippered polygon meshes from range images. In: Proc. SIGGRAPH '94. pp. 311–318. ACM, Orlando, Florida, USA (1994)
20. Vlasic, D., Brand, M., Pfister, H., Popović, J.: Face transfer with multilinear models. In: Proc. SIGGRAPH '05 Papers. pp. 426–433. ACM, Los Angeles, California, USA (2005)
21. Weise, T., Li, H., Van Gool, L., Pauly, M.: Face/off: live facial puppetry. In: Proc. SCA '09. pp. 7–16. ACM, New Orleans, Louisiana (2009)
22. Zhang, L., Snavely, N., Curless, B., Seitz, S.M.: Spacetime faces: high resolution capture for modeling and animation. In: Proc. SIGGRAPH '04 Papers. pp. 548–558. ACM, Los Angeles, California, USA (2004)

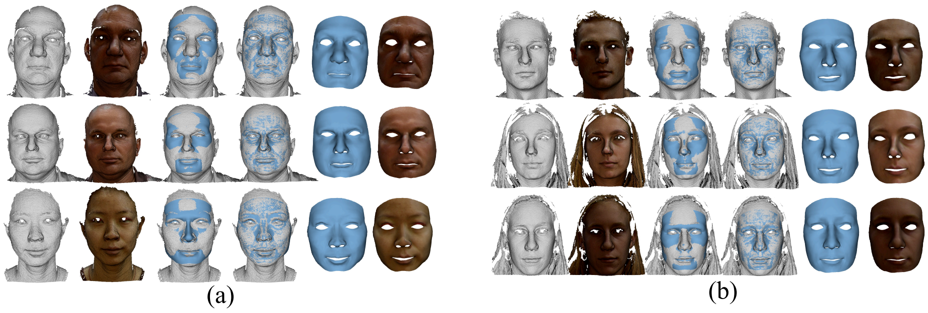

Fig. 4. The results of automatic 3D facial model and texture reconstruction. The six range scans, shown in shaded and texture-mapped renderings in the first and second columns, are from the Face Recognition Grand Challenge (FRGC ver2.0) data set [13]. The third (fourth) column in (a) and (b) shows the overlap between the range scans (gray) and the rigid (non-rigid) registered template model (blue). The final reconstructed facial models are shown in the last two columns in shaded and texture-mapped renderings. All these reconstructed models have the same mesh structures.

Fig. 5. The results of expression transfer. Five facial expressions (anger, laughing, pleased, rage, sad) of the template facial model, shown in the first row, are transferred onto three reconstructed facial models from range scans by the deformation transfer technique.